CCSBT-ESC/2208/09
(ESC Agenda item 7.1)

Australia's National
Science Agency

Update on the SBT close-kin tissue sampling, processing and kin-finding 2022

Jessica Farley, Paige Eveson, Rasanthi Gunasekera

CCSBT-ESC/2208/09

Prepared for the CCSBT Extended Scientific Committee for the Twenty Seventh Meeting of the Scientific Committee

29 August – 6 September 2022

Acknowledgments

The SBT sampling program in Indonesia would not be possible without the significant cooperation of the scientist team at Research Institute for Tuna Fisheries (Bali), and in particular that of Mr Kiroan Siregar and Mr Rusjas Mashar and other staff involved in measuring SBT and collecting tissue (and otolith) samples in Indonesia. The cooperation of the longline tuna industry (coordinated through Asosiasi Tuna Longline Indonesia) and the individual processing companies in providing access and facilities to carry out the sampling is much appreciated. We also thank Adam Kemp at Seatec Pty Ltd for sampling juveniles in Australia, and the Australian Southern Bluefin Tuna Industry Association (ASBTIA) for their ongoing support. We also thank Shane Baylis for work on the R package used for kin-finding ('kinference'). This work was funded by the Commission for the Conservation of Southern Bluefin Tuna and CSIRO Oceans and Atmosphere.

Contents

Abstract

Muscle tissue samples were collected from harvested southern bluefin tuna (SBT) at tuna processors in Port Lincoln, Australia (juveniles; n=1600) in 2022. However, muscle tissue samples were not collected from SBT landed by the Indonesian longline fishery in Bali in 2021/22, due to disruptions caused by Institutional changes in Indonesia. We propose to collect an additional 1500 muscle tissue samples in Indonesia in 2022/23 to compensate for the lack of Indonesian muscle tissue sampling in the 2021/22 season.

Muscle tissue samples collected in 2019/20 and 2020/21 in Indonesia remain in Benoa. Transportation of samples to Australia was prevented due to COVID-19 travel restrictions and delays in renewing the 3-year Implementing Agreement between CSIRO and AMAFRHR for the monitoring program.

Therefore, it was not possible to extract DNA from the tissue for genotype sequencing. We are actively working to have the necessary approvals in place to bring the samples to Australia in September 2022. If successful, we anticipate completing the tissue subsampling, DNA extraction, sequencing, and kin-finding of these samples in time for CCSBT data exchange in 2023.

Given the delay in receiving samples from Indonesia, we were able to complete the tissue subsampling, DNA extraction, sequencing, and kin-finding for both the 2020 and 2021 samples from Australia (juveniles) this year.

The kin-finding analyses to identify parent-offspring pairs (POPs) and half-sibling pairs (HSPs) were updated to include these data and the identified POPs and HSPs provided to the CCSBT in April 2022. A total of 102 POPs and 214 HSPs were identified with high confidence, with a false negative rate of 0.25.To minimise the risk of false positives (i.e., to minimise the number of less-related pairs, in particular half-thiatic pairs (HTPs) incorrectly identified as HSPs), we limited our HSP comparisons to pairs of juveniles born less than 9 years apart. This greatly reduces the number of comparisons between fish that are potentially HTPs (since HTPs are likely to be further apart in age), while not excluding too many potential HSPs. To improve this process, next year we aim to use methods currently being developed using a new genome assembly of SBT that will improve our ability to identify HSPs, reducing the number currently being excluded.

Introduction

In 2013, the Extended Scientific Committee (ESC) developed a new Scientific Research Plan (SRP) for southern bluefin tuna (SBT). The specific projects and priorities for the SRP were considered in both 2014 and 2015. High priority items identified in the work plan included the continued collection and genotyping of tissue samples for close-kin mark-recapture (CKMR) genetics to assess the abundance of adult SBT. The CCSBT has funded the collection and archiving of SBT muscle tissue since 2014/15, and DNA extraction and sequencing of these tissue samples since 2015/16. These samples and data have subsequently contributed to the completion of a second CKMR abundance estimation project that incorporated both parent-offspring pairs (POPs) and half-sibling pairs (HSPs), which was reported to the ESC in 2018 (Davies et al. 2018; 2020). Since

2018, the CCSBT has also funded the analysis of sequencing data to find POPs and HSPs (close-kin identification) on an annual basis. Table 1 shows the work undertaken in each project since 2015. In 2019, the CCSBT agreed to increase the number of tissue samples genotyped from ~2000 to 3,100 annually (i.e., to the number collected annually) to increase the number of "POPs per cohort comparison" (Anon 2019). In this paper we provide an update on progress of activities in 2022.

Table 1. Summary of SBT close-kin work undertaken as part of CCSBT projects each year since 2015. For the genotyping and kin-finding analysis, the season in which the fish were sampled is given.

Project	Muscle tissue collection	DNA extraction & genotyping	Kin-finding	ESC paper
2015	2014/15	NA[1]	NA[1]	CCSBT-ESC/1509/15
2016	2015/16	2014/15	NA[1]	CCSBT-ESC/1609/08
2017	2016/17	2015/16	NA[1]	CCSBT-ESC/1708/09
2018	2017/18	2016/17	2015/16	CCSBT-ESC/1809/08
2019	2018/19	2017/18	2016/17	CCSBT-ESC/1909/08
2020	2019/20 Indonesian samples currently in Indonesia	2018/19	2017/18	CCSBT-ESC/2008/07
2021 (ongoing project)	2020/21 Indonesian samples currently in Indonesia	2019/20	2018/19	CCSBT-ESC/2009/06
2022 (Current project)	2021/22 No sampling in Indonesia	2020/21 Juveniles only & included in kin-finding	2019/20 Juveniles only	Current paper

[1] Genotyping and kin-finding undertaken in FRDC project 2016-044 (see Bravington et al. 2017; Davies et al. 2018).

Muscle tissue collection

Targeted sampling of SBT did not occur at the Benoa Fishing Port in Indonesia in the 2021/22 spawning season due to disruptions caused by institutional changes in Indonesia. We propose to collect an additional 1500 muscle tissue samples in Indonesia in 2022/23 to compensate for the lack of Indonesian muscle tissue sampling in the 2021/22 season.

Note that muscle tissue samples collected in 2019/20 and 2020/21 in Indonesia are currently in Benoa. Transportation of these samples to Australia was prevented due to COVID-19 travel restrictions and delays in renewing the 3-year Implementing Agreement between CSIRO and AMAFRHR for the monitoring program. We are actively working to have the necessary approvals in place to bring the samples to Australia in September 2022.

In Australia, muscle tissue samples were collected from juvenile SBT at the tuna processors during harvest operations in Port Lincoln, South Australia in June-July 2022. Tissue was obtained from 1600 fish ranging from 98-109 cm fork length (FL) to best ensure 3-year-olds are being sampled. The tissue samples were frozen according to protocols provided by CSIRO and have been transported to Hobart. The frozen muscle tissue samples are stored in consecutively labelled boxes with 100 positions (10 by 10) in each box (A01 through J10). Individual samples are given a unique identification label (e.g., SbPL2014_Bx01_A01).

DNA extraction and sequencing

As noted above, muscle tissue samples (adult) collected in Indonesia in 2019/20 and 2020/21 are currently in Benoa. Therefore, it was not possible to extract DNA from the tissue for genotype sequencing as planned. However, all muscle samples collected in 2020 and 2021 in Port Lincoln (juveniles) were subsampled and DNA extracted. The DNA was extracted using a magnetic bead-based extraction protocol (Machery Nagel Nucleomag) kit on an Eppendorf EP motion robot to produce a 90uL archive and 30uL working stock of DNA in micro-titre format plates. Working stock plates of extracted DNA were shipped to Diversity Arrays Technology (DArT) in Canberra for sequencing (referred to as "DArTcap") of approximately 2000 single nucleotide polymorphic loci (SNPs). Archive plates of extracted DNA are stored in a dedicated -80°C freezer located at CSIRO Hobart. All sequencing data (2 years) were sent to CSIRO Hobart in early 2022 for inclusion in the kin-finding (below).

We anticipate receiving the 2019/20 and 2020/21 tissue samples from Indonesia in September 2022. If achieved, we aim to complete the tissue subsampling, DNA extraction, sequencing and kin-finding in time for CCSBT data exchange in 2023.

Kin-finding

The kin-finding analysis database used for identification of POPs and HSPs was updated to include the DArT sequencing data for juvenile (Port Lincoln) samples collected in 2020 and 2021.

Prior to kin-finding, the sequencing data are used to "call the genotype" for each fish and locus in the data (i.e., to infer the pair of alleles present). This genotype-calling entails quite complicated algorithms developed by CSIRO specifically for DArTcap sequencing data which estimates the genotyping error rates for each locus which is important in the identification of HSPs. A plate-level standardization was applied to the sequence count data from all years before calling the genotypes (Farley et al. 2019). This ensured that, for a given loci, the average count across all samples on a plate was the same for each plate.

A series of quality control (QC) steps were applied to the genotyped data to remove fish with unreliable genotype calls. These include:

- a test for heterogeneity to remove fish with an unexpectedly high number of heterozygous loci, which could be an indication of cross-contamination of DNA between individuals;

- a test of whether an individual's genotype could plausibly have been drawn from the 'stock' represented by the rest of the samples to remove fish potentially mis-identified as SBT; and
- a test for an over-representation of null alleles in each individual genotype to remove degraded samples.

After applying the QC steps, 8,409 adults and 18,157 juveniles remained for kin-finding, noting that only juveniles are used in identifying HSPs (Table 2).

POP-finding

Even though we do not have any new sequencing data from adults to include in the analysis, additional POPs can still be identified between the new juvenile data and previously obtained adult data.

We used the genotype data to identify POPs using the same method as the previous two years, which is a modified Mendelian-exclusion statistic referred to as the Weighted-PSeudo-EXclusion (WPSEX) statistic (see Appendix B of Bravington et al. 2017).

Figure 2 shows part of the histogram of the WPSEX statistic comparing across all possible genotyped adult-juvenile pairs (18,157 juveniles x 8,409 adults = 152.6 million comparisons). The POPs are visible as a small bump on the left side and are clearly separated from non-POPs. Most of the histogram (to the right) has been truncated, because otherwise the POPs are too few compared to the gigantic bump of unrelated pairs (the peak of which is around 0.116, where theory predicts it should be based on allele frequencies of each locus) and could not be visualized. The giant bump drops off very quickly to the left of ~0.08, and the flattish tail around 0.05-0.075 will contain a number of adult/juvenile HSPs or grandparent-grandoffspring pairs, which should be somewhat rarer than POPs on demographic grounds.

The number of POPs identified in this data set is 59. Including the POPs that were identified previously using microsatellites (recall that the genotyping method changed after 2015 from using microsatellites to DArTcap sequencing; see Bravington et al. 2015; 2017), we now have a total of 102 pairs. The breakdown by juvenile birth year and adult capture year is given in Table 3.

Table 2. Number of fish used in the kin-finding analyses this year after quality control checks were applied. For the adults, samples were collected from Indonesia in the fishing season ending in the year shown (i.e., samples collected over the 2005/06 fishing season are referred to as year 2006).

Year	Adults	Juveniles
2006	0	1317
2007	0	1325
2008	0	1356
2009	0	1347
2010	972	1315
2011	958	963
2012	536	876
2013	959	903
2014	922	899
2015	0	953
2016	951	854
2017	971	948
2018	700	756
2019	1440	1449
2020	-	1512
2021	-	1384
Total	**8409**	**18157**

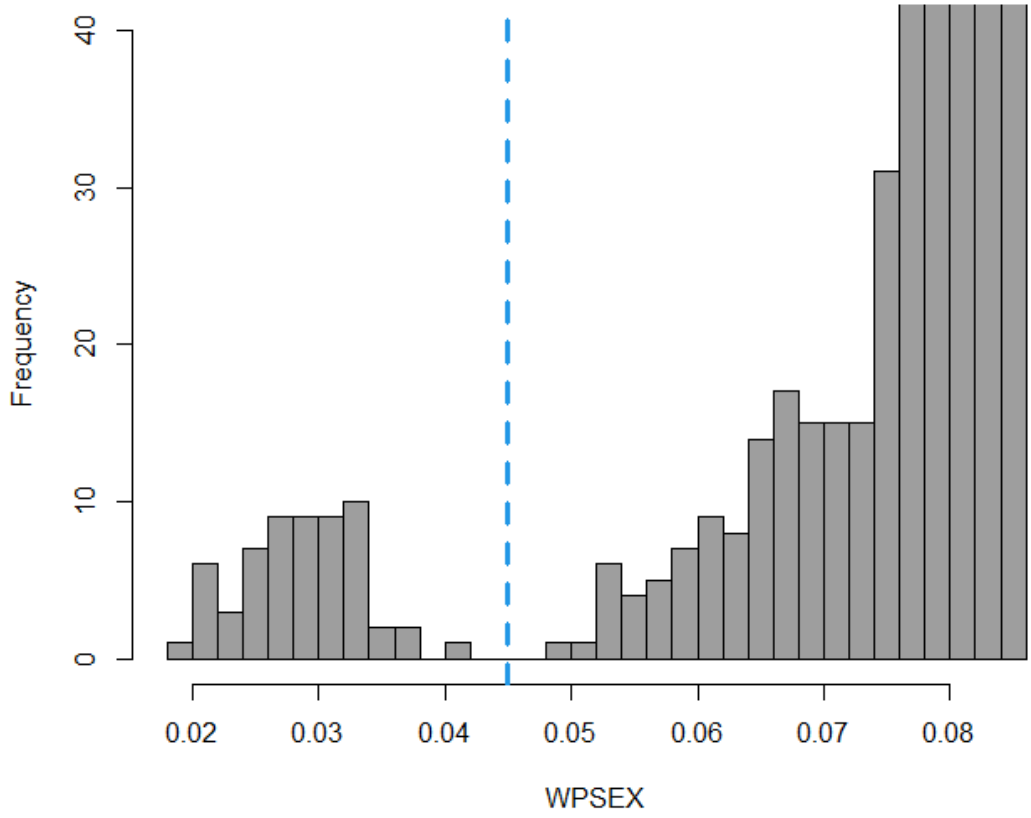

Figure 2. Histogram of the weighted-pseudo-exclusion (WPSEX) statistic for identifying parent-offspring pairs (POPs). Low values (below the vertical blue dashed line) indicate POPs. The x-axis is right-truncated to omit the gigantic peak of unrelated pairs to the right.

Table 3. Number of parent-offspring pairs (POPs) (including those identified using microsatellites and DArTcap data) broken down by juvenile birth year (rows) and adult capture year (columns). NA indicates that no POPs were possible either because no samples exist for that combination of years, or the adult capture year is before the juvenile birth year.

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2016	2017	2018	2019
2002	0	0	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA
2003	0	5	1	2	0	0	0	1	0	1	0	0	0
2004	0	2	0	0	3	0	0	0	0	0	0	0	0
2005	1	4	5	4	1	0	0	1	2	0	0	0	0
2006	NA	4	3	2	0	0	0	0	0	0	0	0	0
2007	NA	NA	3	4	1	3	2	0	2	0	1	0	0
2008	NA	NA	NA	NA	0	1	1	1	0	0	0	2	0
2009	NA	NA	NA	NA	0	1	1	1	0	0	0	0	1
2010	NA	NA	NA	NA	0	0	1	4	0	2	0	0	1
2011	NA	NA	NA	NA	NA	0	1	2	1	2	0	0	0
2012	NA	NA	NA	NA	NA	NA	0	1	1	0	0	1	0
2013	NA	NA	NA	NA	NA	NA	NA	0	0	1	1	3	1
2014	NA	NA	NA	NA	NA	NA	NA	NA	0	0	1	0	0
2015	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	0	0	0
2016	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	2	1	0
2017	NA	NA	NA	NA	0	0	0	0	0	0	3	0	1
2018	NA	NA	NA	NA	0	0	0	0	0	0	0	1	2

HSP-finding

HSPs were again identified using a pseudo-log-odds-ratio (PLOD) statistic to measure the relative probability of a pair of fish having their observed genotypes if they are HSPs compared to if they are unrelated (see Appendix C of Bravington et al. 2017).

When using the PLOD statistic to compare all possible pairs of juveniles (18,157 x 18,157 = 329.7 million comparisons), we do not get a clear separation between the distribution for HSPs and that for unrelated/less related fish, in particular half-thiatic pairs (HTPs). This has been noted as a problem in the past several years, and the overlap keeps increasing as the number of comparisons increases. We dealt with this overlap issue using the same approach as last year (Farley et al. 2021), limiting the comparisons to juveniles born less than 9 years apart (or equivalently, sampled less than 9 years apart since all juveniles are age 3 when sampled). Since HTPs are likely to be greater apart in age than 9 years, this reduces the number of comparisons between fish that are potentially HTPs, thus reducing the size of the problematic HTP bump while not excluding too many potential HSPs. To address this issue long term, a high-quality genome assembly for SBT was developed in 2021 in collaboration with the Wellcome Sanger Institute (UK). New algorithms are being developed that leverage linkage information between genetic markers gained from the genome assembly to improve the accuracy of kin-finding. These algorithms were progressed significantly in the previous year, but still require further testing before applying to SBT.

The PLOD statistic for comparisons between juveniles born less than 9 years apart is shown in Figure 3. The four pairs with PLOD values to the right of 150 are full sibling pairs (FSPs)[1]. The division between HSPs and less related pairs is still not as clear as we would like however, setting a PLOD cut-off value of 50 for HSPs results in only 3 expected false positive HSPs and 214 pairs that we are quite confident are HSPs. Note that we used the theoretical means and approximate variances of the PLOD distributions for HSPs and unrelated/less related pairs (calculated as part of the kin-finding software) to determine the expected number of false positives. The breakdown in numbers of identified HSPs by birth year is given in Table 4.

An inevitable consequence of ensuring that false positives are rare is that a reasonable number of false negatives will be present. Using the expected PLOD distribution for HSPs, we estimated the true number of HSPs to be about 25% higher than 214 because of false negatives. The false-negative rate is allowed for in the population modelling, so is not a problem as long as we have a good estimate of it (Bravington et al. 2017).

[1] Note that all four FSPs were within-cohort pairs, as one would expect for a large adult stock.

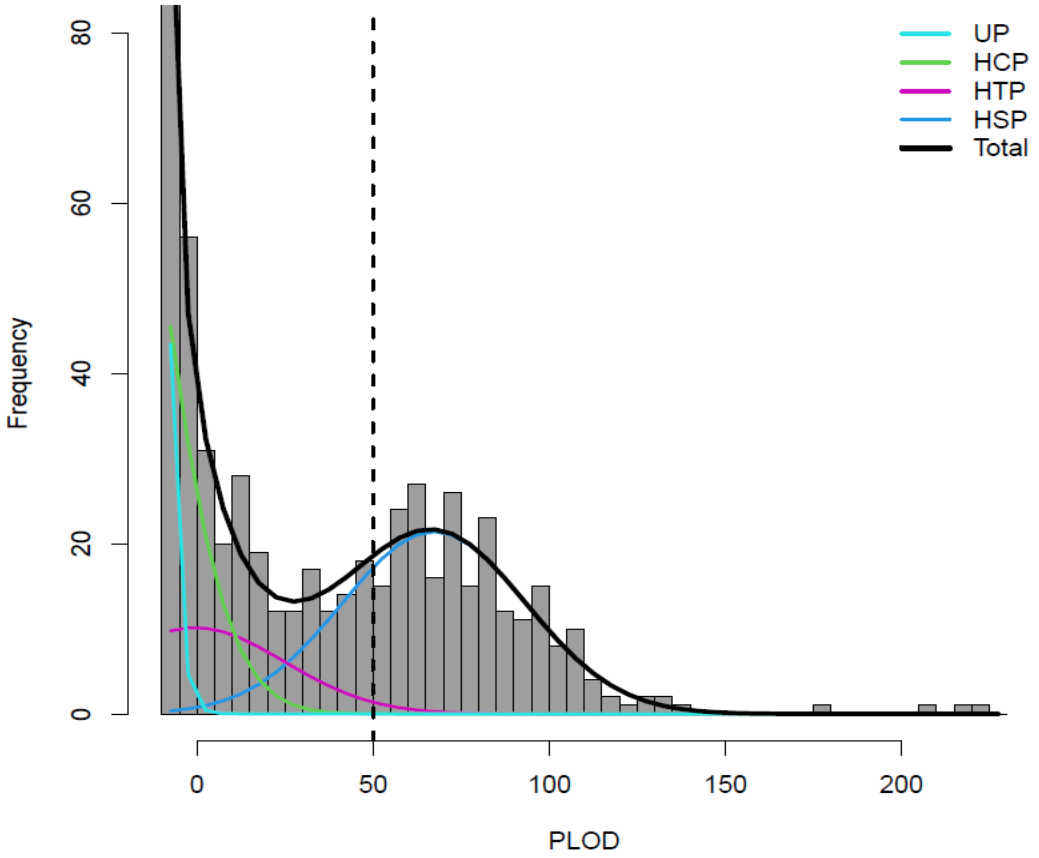

Figure 3. Histogram of the pseudo-log-odds-ratio (PLOD) statistic for pairwise comparisons of juveniles born less than 9 years apart. The approximate PLOD distributions for unrelated (UP), half-cousin (HCP), half-thiatic (HTP) and half-sibling (HSP) pairs are shown. With a lower PLOD cut-off value of 50 for HSPs, we expect ~3 false-positive HTPs. Note that the x-axis is left-truncated to omit the gigantic peak of UPs to the left.

Table 4. Number of half-sibling pairs (HSPs) broken down by birth year of younger sibling (rows) and older sibling (columns). Note that comparisons were only made between juveniles born less than 9 years apart.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
2003	2	4	2	1	0	0	1	0	0							
2004		6	3	6	2	2	1	0	0	2						
2005			5	3	3	3	0	5	1	1	0					
2006				8	4	1	3	5	3	0	1	1				
2007					3	3	2	2	2	2	2	1	2			
2008						5	1	1	2	3	0	1	0	2		
2009							1	2	1	0	0	0	0	4	2	
2010								2	1	2	1	0	1	1	0	0
2011									3	2	1	0	3	4	1	1
2012										3	2	1	1	1	2	1
2013											2	4	1	0	1	0
2014												2	2	1	4	3
2015													4	2	1	3
2016														6	4	5
2017															5	2
2018																5

Summary

The project successfully completed:

1) the 2022 tissue sampling in Australia (juveniles) but unfortunately no sampling was undertaken in Indonesia in 2021/22 (adults) due to disruptions caused by the Institutional changes in Indonesia;
2) tissue subsampling and DNA extraction and sequencing of Australian samples collected in 2020 and 2021 (juveniles); and
3) kin-finding (POPs and HSPs) including the 2020 and 2021 juveniles.

An updated dataset of identified SBT POPs and HSPs was provided to the CCSBT in April 2022. To date, a total of 102 POPs and 214 high confidence HSPs have been identified with the false negative rate for HSPs estimated to be 0.25. As noted in past reports, the overlap between true HSPs and less related pairs, in particular HTPs, continues to increase as the total sample size increases. New algorithms are being developed that make use of a recently generated high-quality genome assembly for SBT to improve the separation and "reclaim" some of the HSPs currently being excluded. While good progress was made on these algorithms this year, further testing is still required prior to application for SBT. Thus, in order to keep the risk of false positives very low, we used the same approach as last year and limited the HSP comparisons to juveniles born less than 9 years apart (see HSP-finding section for details). While this is an adequate solution in the short term, our goal is to apply the new algorithms in our analysis next year.

If 2019/20 and 2020/21 muscle tissue samples arrive from Indonesia by September 2022, we anticipate completing the tissue subsampling, DNA extraction, sequencing, and kin-finding using those samples in time for CCSBT data exchange in 2023.

References

Anonymous. 2019. Report of the Twenty Sixth Annual Meeting of the Commission, Commission for the Conservation of Southern Bluefin Tuna, 17 October 2019, Cape Town, South Africa.

Bravington M, Eveson P, Grewe P, Davies C. 2015. SBT Close-Kin Mark-Recapture: options for the medium term. CCSBT-ESC/1509/19.

Bravington MV, Eveson JP, Grewe PM, Davies CR. 2017. SBT close-kin mark-recapture with parent-offspring and half-sibling pairs: update on genotyping, kin-finding and model development. CCSBT-ESC/1709/12.

Davies CR, Bravington M, Grewe PM, Eveson JP. 2018. Close-kin project report. CCSBT-ESC/1809/14.

Davies C, Bravington M, Eveson P, Lansdell M, Aulich J, Grewe P. 2020. Next-generation Close-kin Mark Recapture: Using SNPs to identify half-sibling pairs in Southern Bluefin Tuna and estimate abundance, mortality and selectivity. Final Report, FRDC Project No 2016-044.

Farley J, Eveson P, Bravington M, Aulich J, Grewe P. 2019. Update on the SBT close-kin tissue sampling, processing, kin finding and long-term sample storage. CCSBT-ESC/1909/08.

Farley J, Eveson P, Gunasekera R, Grewe P, Bravington M. 2021. Update on the SBT close-kin tissue sampling, processing and kin finding 2021. CCSBT-ESC/2108/06.

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

Contact us

1300 363 400
+61 3 9545 2176
csiroenquiries@csiro.au
www.csiro.au

For further information

Oceans and Atmosphere
Jessica Farley
+61 6 6232 5189
jessica.farley@csiro.au
www.csiro.au

Oceans & Atmosphere
Paige Eveson
+61 3 6232 5015
paige.eveson@csiro.au
www.csiro.au